



Hardware

1



Jedním z úkolů práce bylo ukázat, že je možné pomocí veřejně dostupných technologií vytvořit efektivní zařízení. Pro tento a další účely byl v práci testován a použit následující hardware:

1. Raspberry Pi 4 model B
2. ReSpeaker 4-Mic Array  **mikrofon**
3. Krokový motor NEMA17 (47 mm) model 42HD6021-03
4. Světelné senzory TSL2591 v počtu 4 kusů
5. Solární panel

1.1 Raspberry Pi 4 model B

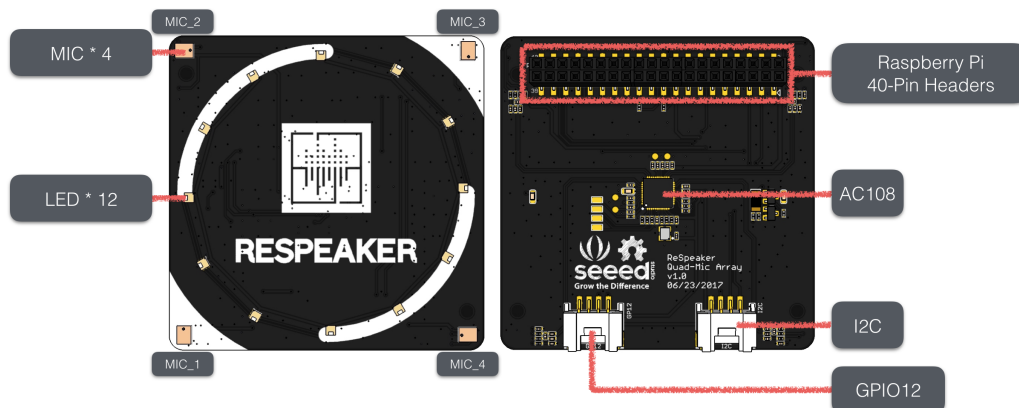
Nízkonákladový a malý jednodeskový počítač  s dostatečným výkonem pro multifunkční úlohy. **Na zařízení byl použit operační systém Linux.** V testovací verzi a v práci byl použit model s novějším typem operační paměti LPDDR4 a 2 GB RAM. Zařízení je vybaveno čtyřjádrovým procesorem ARM Cortex-A72 s taktem 1,5 GHz. Grafika VideoCore s verzí VI. Což dohromady zajišťuje dobrý výkon a vysokou rychlost počítače¹.

1.2 ReSpeaker 4-Mic Array

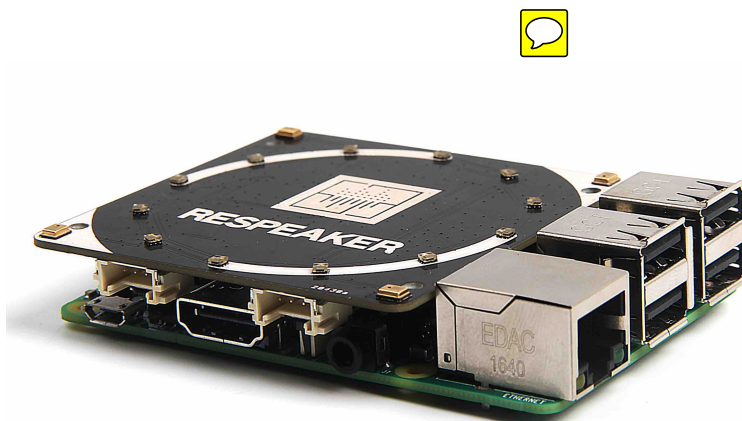
 Je deska vybavená čtyřmi analogovými mikrofony a audio kodekem AC108 pro snímání a zpracování hlasu ve vysokém rozlišení. Zařízení je určeno k použití jako hlasové uživatelské rozhraní **pro řadu počítačů**  Raspberry Pi. Tato verze mikrofону je navíc vybavena LED kroužkem s 12 programovatelnými LED diodami APA102.

¹ <https://rpishop.cz/raspberry-pi-4/1599-raspberry-pi-4-model-b-2gb-ram-765756931175.html>

Díky těmto mikrofonom a LED kroužku má ReSpeaker 4-Mic Array schopnost detekovat hlasovou aktivitu (VAD²), odhadovat směr zvuku (DOA³) a provádět detekci klíčových slov (KWS⁴) široce používaných v aplikacích hlasové interakce⁵.



Obrázek 1.1: Přehled mikrofonu ReSpeaker 4-Mic Array použitého v práci



Obrázek 1.2: ReSpeaker 4-Mic Array mikrofón propojený s Raspberry Pi 4B

1.3 NEMA17 model 42HD6021-03

Malý dvoufázový krokový motor na svůj rozměr s dostatečně velkým statickým momentem 500 mNm a vysokou přesností 200 kroků na otáčku, což je 1,8°. Motor

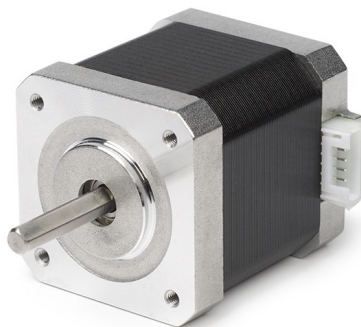
²Voice Activity Detection (VAD) - detekce přítomnosti nebo nepřítomnosti lidské řeči ve vstupním akustickém signálu.

³Direction Of Arrival (DOA) - směr, ze kterého se obvykle šíří zvukové vlny.

⁴Keyword Spotting (KWS) - proces rozpoznávání předem definovaných slov z řečového signálu.

⁵https://wiki.seeedstudio.com/ReSpeaker_4_Mic_Array_for_Raspberry_Pi/

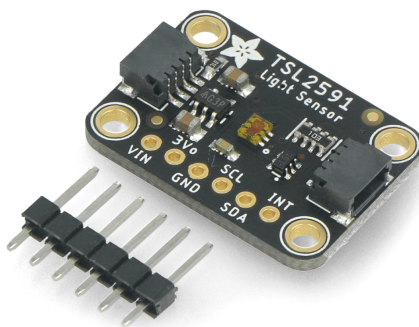
se jednoduše instaluje a demontuje bez použití speciálních nářadí. Často se používá pro 3D tiskárny. V bakalářské práci byl použit k rotaci horní části cílového zařízení.



Obrázek 1.3: Krokový motor NEMA17

1.4 TSL2591

Světelný senzor s vysokým dynamickým rozsahem 600M:1 a velkou citlivostí s účinným maximem do 88.000 Lx. Obsahuje infračervené a celospektrální diody, díky čemuž umožňuje měřit odděleně různé typy světla. Sensor lze snadno spojit s počítačem Raspberry Pi přes I2C rozhraní. V práci světelné senzory slouží k určení místa s nejvyšším dopadem slunečního světla⁷.



Obrázek 1.4: Světelný senzor TSL2591

⁶<https://novo3d.in/product/nema17-stepper/>

⁷<https://rpishop.cz/adafruit/269-svetelny-senzor-tsl2591.html>

1.5 Solární panel

Polykrystalický solární panel ve velikosti 110 x 60 x 2,5 mm. S výstupním napětím 6 V a s výstupním proudem do 200 mA. Solární panel byl použit k zajištění dodatečného nabíjení cílového zařízení⁸.

⁸<https://dratek.cz/arduino/1589-solarni-panel-6v-1w-az-200ma.html>

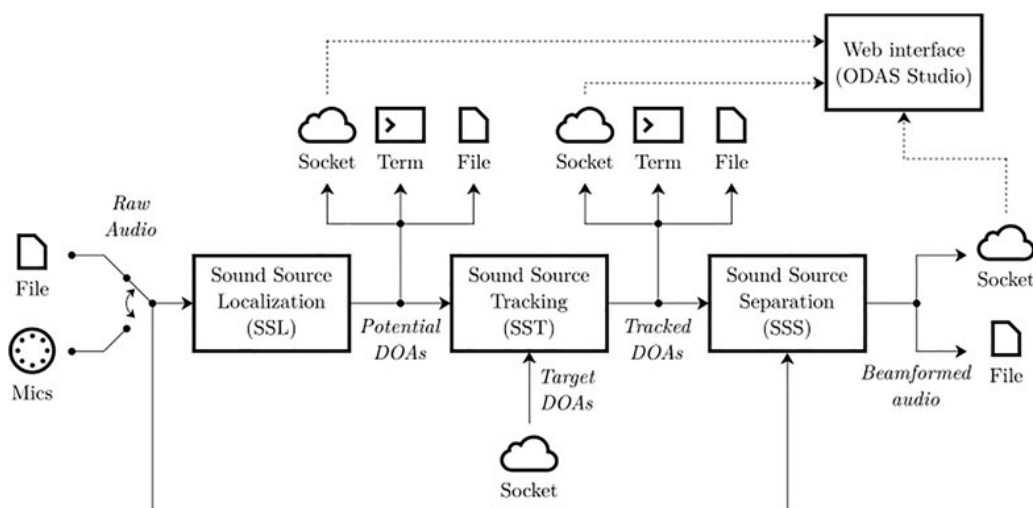
V této části budou podrobně popsány technologie a způsoby jejich aplikace, použité v bakalářské práci. Technologie byly zvoleny na základě poměru jejich snadné integrace a efektivity jejich použití. Některé z technologií byly otestovány a dále využity prostřednictvím realizovaných demo verzí, některé z nich posloužily jako podklad pro vypracování této práce. Základní technologie a nástroje:

1. ODAS (Open embeddeD Audition System)
2. Picovoice - byl použit k detekci hlasové aktivity a potlačení šumu
3. Porcupine - byl použit pro detekci klíčového slova a jeho natrénování
4. ROS - používá se ke spojení všech technologií a jejich komunikace v cílovém zařízení
5. Tkinter - drobný popis...

Vše bylo implementováno v programovacím jazyce Python, byl použit operační systém Linux. Pro testování a modelování byl použit také operační systém Windows.

2.1 ODAS

Open embeddeD Audition System je open source knihovna určená k lokalizaci, sledování, separaci a následnému filtrování zdrojů zvuku. ODAS je framework vyvinutý pro použití na nízkonákladovém a nízkoúčinném vestavěném hardwaru. Je napsaný výhradně v programovacím jazyce C, díky čemuž má lepší optimalizaci a je lépe přenositelný na různá vývojová prostředí [1]. Maximální přenositelnost frameworku je dosažena tím, že má jenom jednu externí závislost na známé knihovně třetí strany FFTW3, která se používá k efektivnímu výpočtu rychlých Fourierových transformací [3]. V rámci práce ODAS byl použit k lokalizaci zdroje zvuku a jeho postfiltraci.



Obrázek 2.1: ODAS schéma zpracování příchozího zvuku

Na obr. 2.1 je znázorněn zvukový kanál pro zpracování zvuku a vstupně-výstupní rozhraní, z nichž každé běží v samostatném vlákne, aby bylo možné plně využít možností vícejádrových procesorů. Zdrojový zvuk může být reprezentován předem nahraným vícekanálovým audio souborem RAW nebo může být získán přímo z připojeného mikrofону pro zpracování v reálném čase. Modul lokalizace zdroje zvuku (SSL¹) generuje pevný počet potenciálních zdrojů zvuku, které se předávají modulu sledování zdroje zvuku (SST²). Modul SST identifikuje sledované zdroje a přenáší tyto DOA do modulu provádějícího separaci zdrojů zvuku (SSS³) k provedení formování paprsku pro každý cílový zdroj zvuku. Uživatel může rovnou definovat cílové DOA, pokud je směr zdroje zvuku předem známý a není vyžadována lokalizace ani sledování. Vygenerované segmenty lze zapsat do audio souborů RAW nebo také odeslat prostřednictvím socketu [3].

2.1.1 Sound Source Localization

Lokalizace zvuku ve frameworku se opírá o metodu Generalized Cross-Correlation with Phase Transform (GCC-PHAT), která se vypočítá pro každý pár mikrofónů. ODAS používá Inverse Fast Fourier Transform (IFFT) algoritmus k efektivnímu výpočtu křížové korelace ze signálů ve frekvenční oblasti. Při práci s malými nebo krátkými poli signálů, ODAS může také interpolovat křížovou korelaci, aby se zlepšila přesnost lokalizace a eliminoval artefakt diskretizace TDOA, který vzniká

¹ Sound Source Localization (SSL) - schopnost určit polohu nebo zdroj detekovaného zvuku podle směru a vzdálenosti.

² Sound Source Tracking (SST) - schopnost sledovat měnící se zdroj zvuku v reálném čase.

³ Sound Source Separation (SSS) - schopnost oddělit aktivní řeč nebo určitý signál od šumu na pozadí nebo ticha.

při Fourierově transformaci. Hlavním principem výpočtu DOA je metoda Steered-Response Power with Phase Transform (SRP-PHAT). Pro každý DOA framework vypočítá SRP-PHAT součtem hodnoty křížové korelace spojené s odpovídajícím časovým rozdílem (TDOA), získaným pomocí GCC-PHAT pro každý pár mikrofónů a vrátí hodnotu DOA s nejvyšším výkonem.

2.2 Picovoice

2.3 Porcupine

2.4 ROS

2.5 Tkinter